

NII Today

89
Sep. 2020

National Institute of Informatics News

Feature

COVID-19と向き合う 情報学の挑戦

データから新型コロナウイルスをみる

CT画像の解析結果を
感染症対策に役立てる

学会や大学と連携して

COVID-19肺炎の診断支援AIを開発

青木茂樹氏 [日本医学放射線学会 理事長]

佐藤真一 [NII コンテンツ科学研究系 教授／

医療ビッグデータ研究センター センター長]

世界中のCOVID-19情報が俯瞰できるサイトを構築

自然言語処理研究者たちによるコラボレーション

相澤彰子 [NII コンテンツ科学研究系 教授／副所長]

Commentary With コロナ時代の

オープンコラボレーションによる研究

河原大輔氏 [早稲田大学 基幹理工学部 情報通信学科 教授]

「SNSによるデマ拡散」問題の本質とは

「疑うタイミング」の創出が重要に

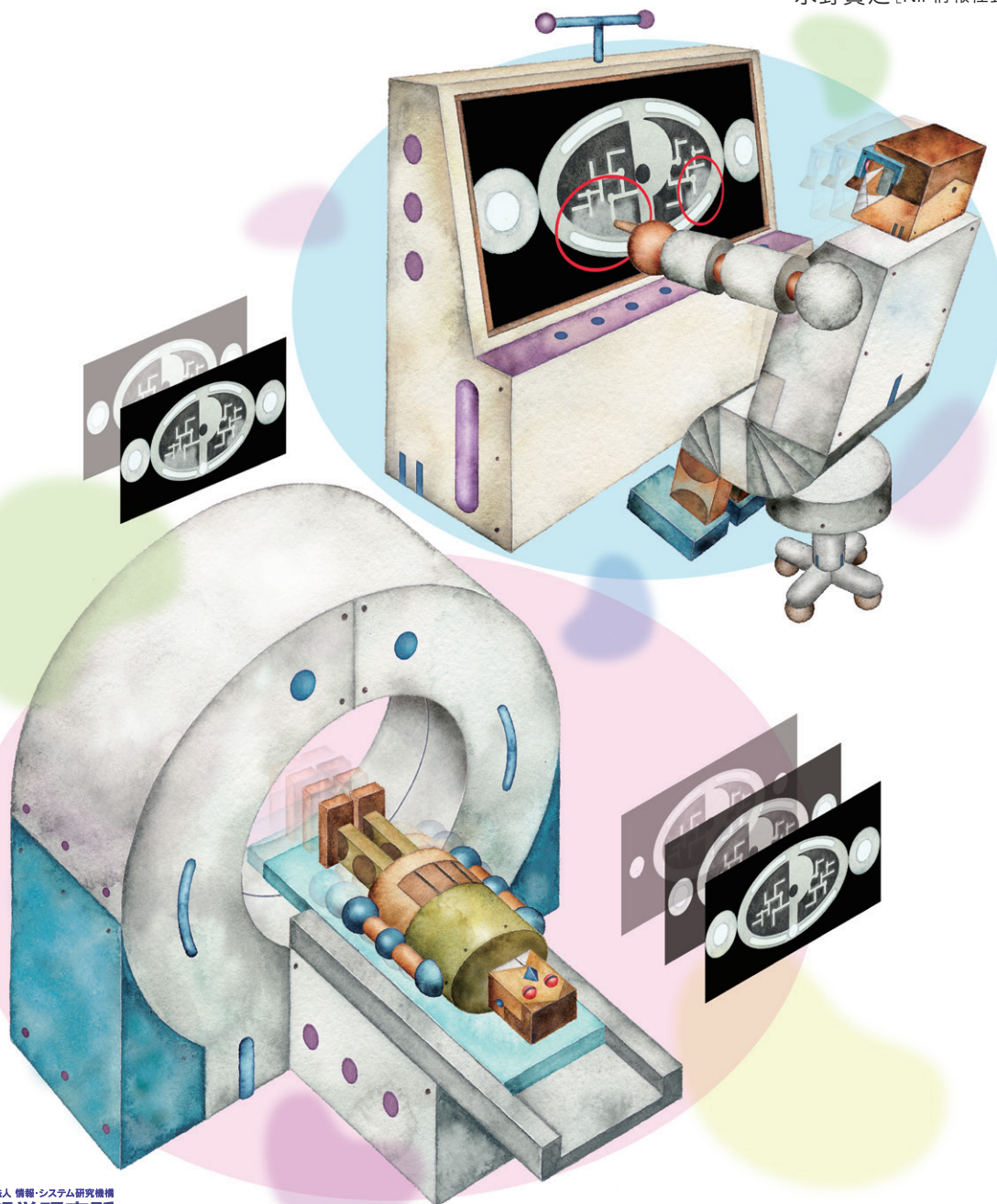
鳥海不二夫氏 [東京大学大学院工学系研究科

システム創成学専攻 准教授]

ビッグデータ解析でコロナ社会の課題解決を

外出自粛率や人の接触頻度を「見える化」する

水野貴之 [NII 情報社会相関研究系 准教授]



CT画像の解析結果を 感染症対策に役立てる

学会や大学と連携してCOVID-19肺炎の診断支援AIを開発

青木茂樹氏 Shigeki Aoki

順天堂大学大学院医学研究科 放射線診断学 教授／
日本医学放射線学会 理事長

佐藤真一 Shin'ichi Satoh

国立情報学研究所 コンテンツ科学研究系 教授／
医療ビッグデータ研究センター センター長

聞き手：田井中麻都佳 Madoka Tainaka

『NII Today』編集デスク／サイエンスライター

現在、COVID-19の診断にはPCR検査が用いられているが、より精密な診断や予後の予測にはCT（コンピュータ断層撮影法）検査が有用だという。とくに日本の医療機関における人口あたりのCT装置保有台数は世界一を誇り、CT画像データの活用に大きな期待が寄せられている。そうしたなか、NIIの「医療ビッグデータ研究センター（Research Center for Medical Bigdata：RCMB）」は日本医学放射線学会とともに、収集したCT画像を用いてCOVID-19肺炎の診断支援AI（人工知能）の開発を進めてきた。その成果とRCMBの取り組み、医療支援AIの可能性などについて、日本医学放射線学会の青木茂樹理事長とRCMBの佐藤真一センター長に聞いた。

より精密な診断や予後予測に役立つCT検査

—— COVID-19の診断に、CT検査はどのように活用されているのでしょうか？

青木 COVID-19の方の肺のCT画像を見ると、多くの場合、ウイルス性肺炎特有のすりガラス状や網目状の影が見られます。さらにそれが1カ所だけでなく、末梢や両肺に見られるのが特徴的です（図1）。たとえば多くの感染者を出したダイヤモンド・プリンセス号では、PCR検査で陽性となった方のうち、112例についてCT検査が実施されました。そのなかで症状があったのは30例で、82例は無症状でしたが、この82例の半数以上の44例の肺にCOVID-19に特徴的とされる所見が見られたのです。CTはX線写真と比べてたいへん感度の高い検査方法なので、微細な病巣の検出に役立つといえます。

—— 無症状の方の半数以上に肺の炎症が見られたのは驚きです。

青木 人間の身体は予備能が高いため、組織の形状が多少変化しても正常範囲内で機能するということがよくあることで

す。沈黙の臓器と言われる肝臓などは、肝硬変でカチカチに硬く小さくなってしまっても、症状が出ないことがあります。つまりCT検査は、無症状の方の病巣を見つけたり、より精密に症状を把握したりするのに有効なのです。

ただし、CT検査をCOVID-19のスクリーニングに使うことは学会としては推奨していません。というのも、CTは感度は高いものの、特異度は比較的低いためです。特異度が低いとは、陰性のものを間違えて陽性と判断してしまう可能性が高いということ。肺線維症やそのほかのウイルス性肺炎と区別することが難しく、感染初期では変化が軽微なこともあって、専門医でないと病変を見つけにくいのです。

もっとも、ほかのウイルス性肺炎が蔓延していない状況であれば、肺に疾患のない患者群に対してCT検査をCOVID-19のスクリーニングに用いることは、ある程度は効果があります。したがって、学会のガイドラインでは、CT検査をPCR検査の代用とすることは推奨しないが、COVID-19の疑いがある患者に対する医療行為の意思決定に利用することは許容される、という微妙な表現をしています。

—— ではどういった方にCT検査を実施するのがいいのでしょうか？

青木 PCR検査が陽性で、症状のある方、悪化傾向にある方は、CT検査を受けたほうがいい。病変を詳細に診るのに役立ちますし、快方に向かっているのか、さらに悪化するか、予後の予測にも有用です。



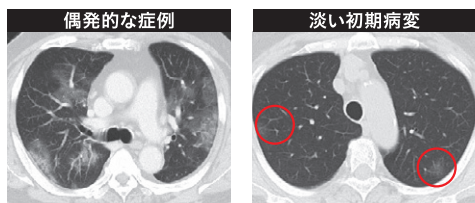
青木茂樹

1959年東京生まれ。1984年 東京大学医学部卒業。臨床研修の後、1987年米国UCSF（カリフォルニア大学サンフランシスコ校）神経放射線部門留学。1995年山梨医科大学放射線部助教授、2000年東京大学大学院医学系研究科放射線医学助教授を経て、2008年より現職。現在、日本医学放射線学会理事長の他、日本磁気共鳴医学会理事長、中央社会保険医療協議会技術評価委員。2012年第107回医師国家試験委員長。



佐藤真一

コロナウイルス感染症での画像診断の有用性



→ COVID-19 肺炎疑いと診断
→ その後、PCR 陽性と判明

日医放 J-MID

図1 左の写真は症状もなく、偶発的に見つかったCOVID-19肺炎の症例。特徴的なCTの所見から感染を疑い、その後PCR検査で陽性と判明した。CTはCOVID-19の診断に有用だといえる。ただし、CTでは病変を指摘できない症例や、右の写真の赤丸で示すように感染初期では変化が軽微で専門家でないと病変を指摘しにくい場合がある。このような人間では見つけにくい病変をAIが判別することが期待されている。(画像提供: 順天堂大学医学部放射線医学教室・放射線診断学講座 明石敏昭 准教授)

画像診断医の診断を教師データに用いてAIの精度を高める

—— 日本医学放射線学会とRCMBは、共同でCOVID-19の診断支援AIの開発をされました。経緯を教えてください。

青木 学会では5年ほど前から、東京大学、京都大学、大阪大学、岡山大学、九州大学、慶應義塾大学、順天堂大学からすべてのCT画像を集め、J-MID (Japan Medical Imaging Database) データセンターに蓄積するプロジェクトを進めてきました。一昨年末から本格的に収集できるようになり、そのデータをNIIのRCMBのクラウド基盤に送付して、すでに1億3000万枚以上のレポート付きCT画像によるビッグデータを構築しています。これらを活用して診断支援AIの開発を進めてきた経緯があり、COVID-19対策にも役立てたいと考えました。

佐藤 RCMBは、医療ビッグデータクラウド基盤の構築とAIによる医療画像解析研究のハブとしての役割を担うため、日本医療研究開発機構 (AMED) の予算支援のもとに2017年にNII内に設立された組織です。特徴的なのは国内の主要な研究機関と連携して研究チームを編成している点で、機械学習や深層学習、ネットワーク、クラウド、セキュリティ分野などの研究者とともにさまざまな解析を行っています。また、解析に必要な不可欠なリアルな医療画像の収集については、日本医学放射線学会をはじめ、現在6つの医学系学会に提供していただいています。これらの画像を匿名化したうえで、NIIの医療画像ビッグデータクラウド基盤に投入して画像解析を進めています。

そして今回、COVID-19拡大の状況を受けて、J-MIDとオンラインで接続されている7施設からすでに128例のCOVID-19のCT画像を提供していただき、日々新たな画像を送信していただいております。また、ダイヤモンド・プリンセス号のCT画像を保有する自衛隊中央病院からは240例を提供していただき、AI開発を進めました。

—— それらが深層学習における教師データとなったわけですね？

佐藤 はい。CT画像に限らず、画像解析に深層学習を用いる際には、たくさんの教師データが必要です。今回は正解データに、実際の患者さんの画像とPCR検査による陽性／陰性の結果、画像診断医による診断結果を用いました。その結果、8割の精度で正しく判定できるAIを開発することができました。

今後さらに症例数が増えれば、より精度を高めることができるでしょう。ただしPCR検査は偽陰性、つまり陽性なのに陰性という結果が出てしまうことが多いため、これを正解データとして採用すると成績が伸びないことがわかりました。そこで現在は画像診断医による診断結果を正解データとしています。

異常データが極端に少ないという不均衡を超えて

—— 症例のデータが400例弱というのは、深層学習に投入するには少ないように感じます。それでも精度は出るものなのですね？

佐藤 教師データには、明らかにCOVID-19肺炎ではない画像も不可欠です。J-MIDには2019年11月以前のデータも大量にありますが、とくに肺に影があるのにCOVID-19でないというデータは、AIを賢くするための難しい訓練データとしてたいへん有用です。症例の画像だけでなく、背景に何万というビッグデータがあればこそ、精度を高めることができるのです。

そもそも医療データには、正常なデータに比べて症例のデータが圧倒的に少ないという特徴があります。そうした不均衡なデータに対して、アルゴリズムをさまざまに工夫して精度を高めるのも、我々の研究の大きなテーマの1つなのです。

—— すでに医療現場で活用され始めているのでしょうか。

青木 いずれ広く活用していきたいと考えていますが、現場で使うためには薬事承認や倫理委員会の承認など、さまざまな手続きを経る必要があります。まずはJ-MIDに参加している大学を中心に、実証実験を通じて精度を高めていきたい。現在、順天堂大学ではCOVID-19の疑いのある方には画像診断医がずっと張り付いて、CT検査の結果を見て、疑いが濃厚なら、PCR検査の結果を待たずに陽性患者と同等の対応をしています。このCTに張り付いている画像診断医の役割をAIで代替できれば、医療現場の負担を大いに軽減できるでしょう。

なお、COVID-19肺炎の画像解析AIの開発は国内外で急ピッチに進められており、すでに薬事承認を受けたものもあります。ただしいずれも精度はさほど高くありません。コロナ禍で素早く薬事承認が下りるようになったのはいいことですが、医療現場で安心して使っていくためには、まだまだ検証を重ねる必要があります。

ビッグデータから感染症の初期の流行を捉える

—— COVID-19に関して、放射線学会ではほかに「ウイルス性肺炎サーベイランス」というユニークな取り組みをされていますね。

青木 これは放射線科医がCT検査でウイルス性肺炎の症例を見たら、COVID-19らしさを評価したうえで学会のWebサイトに報告するという試みです。画像を送るのではなく、所見を報告するだけの簡単なシステムで、すでに1300人分のデータが集まりました。その結果を棒グラフにして時系列で示したところ、まったく別の検査かつ別のデータ収集法であるはずのPCRの陽性者数 (厚生労働省発表) とグラフの形が重なりました。これを見れば、第二波が収束に向かいつつあることや、第一波に比べて第二波のほうが比較的症状が軽い人が多いことな

どがわかります。

さらに特徴的なのが、第一波の初期に、PCR検査による陽性者数に先行してウイルス性肺炎の患者が増えている点です。つまり、このシステムが未知の感染症の早期の検知に役立つ可能性を示唆しています（図2）。

現在はデータを人の手で入力していますが、今後、NIIと協力してクラウドにCT画像を投入するだけで、コロナ肺炎らしさなどを自動で判定・入力できるようにしたいと考えています。また、流行の全体像を把握するためには、J-MIDに接続している7大学だけでなく、感染症指定医療機関などの協力も得てデータ数を増やしていく必要があります。実現すれば、未知の感染症の検知やクラスターの拡大を初期段階で食い止めるのにも効果を発揮するはずです。

医療の支援に役立つ医療画像ビッグデータの可能性

—— RCMBでは、COVID-19以外にもさまざまな医療画像の解析に取り組んできました。

佐藤 日本医学放射線学会とは、CT画像を用いた、くも膜下出血の検出や腎臓がんの検出、造影剤を用いることなく血管を推定して見える化するAIの開発などを進めています。ユニークなものに、MRI（磁気共鳴画像診断装置）による脳腫瘍画像に関する研究があります。これは、正常なデータに比べて病変のデータが極端に少ないことから、「敵対的生成ネットワーク＝GAN（Generative Adversarial Networks）」というフェイク画像の生成などに使われる深層学習の手法を使って、本物そっくりの病変データを作成して病変画像を増やし、AIの精度を高める研究です。そのほか、大量の骨盤のCT画像を用いて、年齢や性別によって日本人の骨盤がどう変化していくのか、腰痛と関連のある骨盤後傾について調べるなど、解剖学の新しい知見を得るためのチャレンジもしています。

ちなみに、最初に着手したくも膜下出血の検出はいまだに開発の途中です。くも膜下出血は素人目には出血がまったくわからない画像もあって、見た目のバリエーションがとても多いた

インタビュアーからのひとこと

第三次AIブームに火をつけた深層学習のなかで、飛躍的な成果をあげたのが「画像解析」だ。条件さえ整えば、いまやエラー率は人間より低く、とりわけ医療画像解析AIは専門医の診断支援に役立つとして大いに期待されている。ただし、医療データの収集やAI診断の活用が進むかどうかは、受容する側の人間次第である。コロナ禍のなかで、ゼロリスクを追い求めても立ち行かないことを多くの人が実感したはずだ。人命にかかわることだけに難しい問題だが、間雲にシンギュラリティを恐れるより、AIを優秀な相棒とするために人間の側がリテラシーの向上に努めるほうが得策ではないか。

田井中麻都佳

2012年より『NII Today』編集デスク。中央大学法学部卒。科学技術情報誌『ネイチャーインタフェイス』編集長、文部科学省 科学技術・学術審議会情報科学技術委員会専門委員などを歴任。書籍編集や執筆、大学や企業の広報媒体を中心に活動中。NIIの河原林健一氏との共著『これも数学だった!?～カーナビ、路線図、SNS』（丸善ライブラリー）がある。専門家の言葉をわかりやすく伝えるインタープリターの役割を追求している。



めです。特徴がつかみにくいため、AIの精度が出ないのです。

青木 当初、私たち医療研究者はAIに対して過度な期待を抱いていたんですね（笑）。くも膜下出血は絶対に見逃してはならない病気の1つですが、夜間で専門医がいないと判断が難しくて見逃してしまうこともあります。微妙な症状の所見をAIで拾えるなら、当直医もAIに相談できるし、わざわざ夜間に専門医に聞かなくても、次の処置に進むことができる。実現すればたいへん有用なAIになります。

いずれにせよ、RCMBと共同で開発を進めるなかで何度も話し合いの場をもち、AIで解くべき問題をとともに考え、AIは何かできて、何が不得手なのかを理解できたのは非常によかった。肝臓がんの検査で偶然に腎臓がんを見つけることがあるように、CTやMRIの画像は多種多様な情報をいっぺんに写し撮るのが特徴的で、そこから読み取れる情報はまだまだあるはずです。また、すべてデジタル化されているのでAIに投入しやすい。しかも日本の医療機関は質のいいデータを膨大に保有しています。これらをうまく活用すれば、世界に通用するAIを生み出すことができるでしょう。ぜひ、NIIでさまざまなユースケースを示していただき、さらなるデータ収集

や活用を促していただきたいと思います。

佐藤 やはり現場で役に立つことを示していかなければなりませんね。すでに、病理画像から胃がんを検出するAIを開発し、福島県の医療機関ネットワークで活用されている例もあり、これに続くユースケースを示していきたいと考えています。また、個別の疾患に対応するだけでなく、医療画像のビッグデータを活用することで、先のウイルス性肺炎サーベイランスのように、未知の感染症の流行や疾病を早期に捉えるといった異常検知に役立つこともできるでしょう。今後はさらにデータを増やして、より現場で役立つ医療画像診断支援AIの開発や医療インフラの構築をめざして研究を加速していきたいと思います。

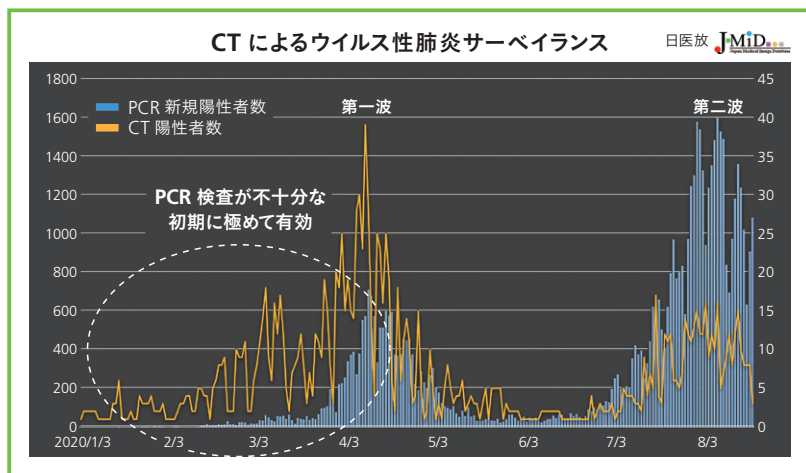


図2 青の棒グラフのPCR陽性の新規患者数と見比べると、黄色の折れ線グラフのCTサーベイランスでは第一波だけでなく、第二波も正確に捉えていることがわかる。検査数が限られているPCR検査で流行の全体像を捉えられているのかは不明だが、簡便・迅速に行うことのできるCT検査は検査数が多く十分に活用できるため、PCR検査体制が不十分な初期に、感染拡大状況を捉える極めて有用な指標になる。

世界中のCOVID-19情報が 俯瞰できるサイトを構築

自然言語処理研究者たちによるコラボレーション

相澤彰子 Akiko Aizawa

国立情報学研究所 コンテンツ科学研究系 教授／副所長
東京大学大学院 情報理工学系研究科コンピュータ科学専攻 教授／
総合研究大学院大学 複合科学研究科 教授

聞き手：濱門麻美子氏 Mamiko Hamakado

岩波書店 自然科学書編集部 編集長

新型コロナウイルス感染症の拡大を受け、自然言語処理の研究者たちが各自の専門性を活かしてCOVID-19に関する世界中のさまざまな情報を集約し、日本語で提示するサイトを短期間で構築した。このコラボレーションはどのように進められたのか。プロジェクトから得られたものは何か。プロジェクトの開始時から参加しているNIIの相澤彰子教授に聞いた。

研究者コミュニティとして取り組んだプロジェクト

——「COVID-19 世界情報集約サイト」とはどんな内容ですか。

相澤 サイトの構造は、「表」と捉えていただくとイメージしやすいでしょう。縦・横に2つの直交する軸（項目）があり、1つの軸は日本／中国／アメリカ／ヨーロッパ／アフリカといった「国・地域」です。もう1つの軸が感染状況／予防・防疫・緩和／医療情報といったトピックの「カテゴリー」です（図1）。サイトに行くと、「国・地域」×「カテゴリー」からなる表の情報を見せる窓が並んでいます。じつは、この表にはさらにもう1つ、時間軸があります。ほぼリアルタイムで情報を収集していて、それぞれの窓のなかで新着順に記事が並んでいます。

—— どのようにしてこのプロジェクトが始まったのですか。

相澤 セっかく自然言語処理の研究者がいるのだから、この状況で何か役に立つことをしたらどうかという、NIIの喜連川優所長からの提案がきっかけです。その提案に賛同した研究者たちが議論を始めて、プロジェクトのメンバーが集まりました。

東日本大震災のときも、若い研究者がボランティアに震災関連の情報収集に取り組む活動はありましたが、世界中の研究グループがこれほど大規模に迅速に動いたのは、自然言語処理分野でも今回が初めてだと思います。分野の中核となる国際会議でも、立て続けにCOVID-19のワークショップが緊急開催され、主催者の予想を超える60本以上の論文が集まりました。

各研究室で役割を分担してサイトを構築

—— サイトの設計はどんなふうに行ったのですか。

相澤 京都大学の黒橋禎夫教授や早稲田大学の河原大輔教授が



相澤彰子



図1 | COVID-19世界情報集約サイト : <https://lotus.kuee.kyoto-u.ac.jp/NLPforCOVID-19/>



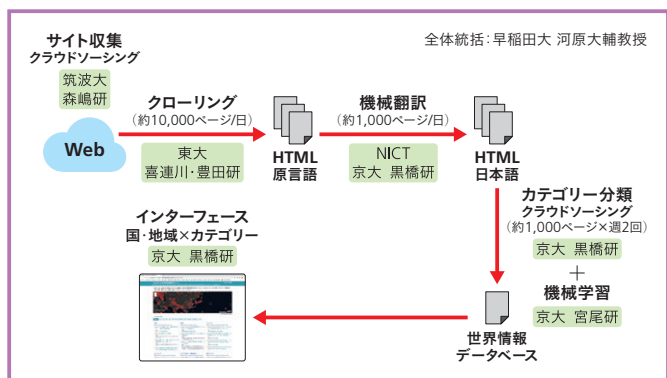


図2 | サイト構築の流れと役割分担 (p.7参照)

ラフなスケッチをつかって、みんなで議論しました(図2)。先ほどお話しした2つの軸として何を設定するか、つまり、どの国から情報を集めて、どういうカテゴリーにまとめ上げるかという点は、試行錯誤しながら決めていった部分です。

—— **サイト構築の流れはどうなっていますか。**

相澤 まず、情報ソースを決めます。信頼できる情報を集めるために、政府系などを中心に、定期的に最新の情報を収集する先のサイトのリストをつくります。このサイトの選定にはクラウドソーシングも使っています。その国・地域のクラウドワーカーが信頼できると確認したサイトを収集先にするのです。

収集先に決まったサイトをクロール(巡回・収集)しますが、クロールにもかなり技術が必要です。サイトの内容は頻繁に更新されるので、毎日その内容をチェックしに行きます。ページ内に埋め込まれた参照コンテンツへの対応も必要です。そうした作業を経て、クロールしてきたページはいろいろな国の言語で書かれているので、機械翻訳にかけてすべて日本語にします。ご存じの通り、昨今の機械翻訳はたいへん性能がよく、文書の内容をつかむには十分です。

もとの言語と日本語がそろったページができたら、「国・地域」ごとの「カテゴリー」に記事を振り分けます。記事数が多いので、現在は機械学習を使って、ほぼ自動的に振り分けています。こうしてできたデータベースを、いま見えているサイトのインターフェースで表示します。

それぞれの要素技術は、このプロジェクトのために考案されたものではありませんが、それらを組み合わせで動かす意義は、単なる $1 + 1 = 2$ の足し算ではありません。IBM が開発した人工知能(AI)である Watson が、2011年にクイズ番組で人間のチャンピオンを破ったときも、ポイントは個々の要素技術ではなく、それらを組み合わせで動かしたところにありました。

サイト拡張の可能性も視野に

—— **今後はどんな展開が考えられますか。**

相澤 いろいろなモジュールを考えて、このサイトに組み込んでいく可能性があります。自由にお試しの機能をテストできることがサイトを手づくりする魅力で、可能性はいくつか話題にのぼっています。たとえば、時系列の解析があったらいいな、とか。このサイトは新しい記事が追加されてどんどん流れてい

くのですが、何月何日頃にという話題がどこで盛り上がっていたかといったことが可視化できると、俯瞰性が高まります。

また、Twitterのようなソーシャルメディアと結びつける、論文などの専門的知識とつなげていく、というも拡張の方向性として考えられます。

プロジェクトの意義と見えた課題

—— **このプロジェクトの意義をどうとらえていますか。**

相澤 「誰かの役に立つことができれば」という気持ちで始まったプロジェクトですが、プロジェクトの参加者にとっても、サイトの構築を通して、世の中ではこういうことが求められているんだとか、いまの自分たちの技術ではこういったところが足りないんだとか、そういう知見が得られることは研究面での価値だと思います。

COVID-19に関しては、米国のアレン人工知能研究所がいろいろな研究所や機関と協力して関連の論文を集めたデータセットを構築しています。そのデータセットを使ったテキストマイニングの研究で浮かび上がった課題の1つが、COVID-19に関する専門用語の辞書がないことです。専門用語を論文から抽出するには、人手をかけてタグ付けをしたアノテーションデータが必要なのですが、それが十分にはないのです。アノテーションデータのような足場がないと、詳細な分析ができません。

生命科学は専門用語辞書やオントロジーが非常によく整備された分野ですが、それでも今回のようなことが起きると、やはり足りないことがわかります。信頼できるアノテーションが効率的にできるような仕組みを、いかに少ないデータでつくっていくかが技術的な課題です。

「COVID-19世界情報集約サイト」の話に戻ると、ここでもアノテーションに関する新鮮な驚きがありました。クラウドソーシングに出せば何でもできると思っていたけれども、言語、地域、カテゴリーが変わると、十分なクラウドワーカーが確保できないことが当たり前になることです。あるドメインにおけるAIシステムの成否が、そのドメインのデータをアノテーションできるクラウドワーカーが確保できるかどうかにかかっていると考えると、この問題は非常に大きなバイアスや格差をもたらし得るものと実感しました。世界中で同時に起きているリアルな社会の膨大なデータを扱って初めて知ったことです。

インタビューからのひとこと

世界中の情報が日本語で俯瞰できるこのようなサイトは、政治・行政やマスコミにまず利用してもらい、目配りの利いた政策や報道に活かしてもらいたいと感じた。

本文中に機械翻訳の精度向上への言及がある。今回の原稿を作成するにあたり、音声を手動でテキスト化するサービスを利用して、実用性のある域に入っていることを知った。機械翻訳に限らず、自然言語処理技術の恩恵をダイレクトに感じることの多い昨今である。



濱門麻美子

岩波書店 自然科学書編集部 編集長
東北大学理学部卒業、東京大学大学院理学系研究科博士前期課程修了、理学修士。1993年、岩波書店入社。自然科学書編集部に配属され、認知科学・言語学・数学などの書籍を担当。2015年より現職。

Withコロナ時代の オープンコラボレーション による研究



河原大輔氏 Daisuke Kawahara

早稲田大学 基幹理工学部 情報通信学科 教授
1997年京都大学工学部卒業。1999年同大学院修士課程修了。2002年同大学院博士課程単位取得認定退学。東京大学学術研究支援員、独立行政法人情報通信研究機構主任研究員、京都大学大学院情報学研究科准教授を経て、2020年より現職。自然言語処理、知識処理の研究に従事。博士（情報学）。

2020年3月上旬、新型コロナウイルス感染症（COVID-19）が中国、韓国、イタリアに続き、日本でも広がりつつあった。そうしたなか、NIIの喜連川所長から、JST（科学技術振興機構）CREST・さがけ複合領域「ビッグデータ基盤」（喜連川総括）の自然言語処理研究者に向けて、COVID-19に対して何かできないかという問いかけがあった。まさに国難、いや「地球難」と言えるような状況にあって、当然、我々としても自然言語処理技術を社会に役に立てたいという思いがあった。こうして、有志の研究者たちの手で、本プロジェクト（p.5～6参照）が立ち上げられることになった。

しかし、コロナ禍では、リアルに顔を突き合わせて議論することができない。そこでプロジェクト当初から、オンライン・ミーティングにはZoomを、日々のコミュニケーションにはSlackを、コードの共有にはGitHubを用いてきた。

初期の頃のSlackのログを見返すと、次のような案が挙がっていたことがわかる。

- ファクトチェック
- 論文分析：COVID-19関連論文からの知識発見
- 各国の母国語で発信される情報の収集、翻訳
- 感染者数などの数値データの言語化、変化点検出
- 時系列イベント解析
- どこで何が問題になっているかの認識

これらの案をもとに議論をするなかで、ファクトチェックと論文分析は他の研究グループで取り組みが始まりつつあった

め、それ以外で役に立つものを考えることにした。こうして、各国の情報（ウェブページ）を収集、機械翻訳、カテゴリーに分類して、国×カテゴリーという2軸で表示する「COVID-19世界情報集約サイト」の構成案を固めていった。

このとき議論の焦点となったのが、サイトのユーザーとして、誰を想定するかだった。官僚・政治家、企業幹部、医療関係者、一般市民など、さまざまな議論を重ねたが、結論が出ないまま、まずはサイトを構築し、使いながら検討することとした。

プロトタイプができたのは3月下旬のことだ。機械翻訳には当初、Google翻訳を使っていたが、大量のページを翻訳するには費用がかかりすぎるため、フリーの「みんなの翻訳」（NICT）に切り替えた。交渉の結果、無制限のアクセスにしてくれたことができたので、非常に助かった。

カテゴリーについては、感染状況、経済対策などの6つを設定し、当初はキーワードに基づく分類器をつくって、データ収集にあたった。しかし、カテゴリー分類および翻訳は自動処理で、精度には限界がある。情報技術の常であるが、ある程度の精度を出さなければ、誰にも使ってもらえない。そこで、5月からカテゴリーをクラウドソーシングで付与することにした。各ページのカテゴリー付与を複数人のクラウドワーカーに依頼し、その結果を集約することで高い精度を得ることができるようになった。しかしこれも費用がかかりすぎることから、結局、2カ月間で終了した。

その後、これまでのカテゴリー付与の

データを訓練データとして、汎用文脈言語モデルBERTを用いて学習することによって分類精度を向上させることができるようになった。さらに、我々で分担して、誤ったカテゴリーを手で修正する仕組みもつくった。最後は人海戦術である。

こうした苦労を経て、6月には本サイトの一般公開に漕ぎつけることができた。それぞれの研究者が得意分野を活かして、役割分担しながら進めたことで、サイトの素早い構築が可能になったのである。なお私は、各研究室から上がってくるモジュールを統合するディレクターの役割を担った。

今後の課題としては、このサイトを必要とするユーザーへ届けること、また、より効率的に役に立つ情報を発見できるようにデザインを改善していくことなどが挙げられる。

今回、カテゴリー分類にBERTを用いたが、BERTの学習はpre-trainingとfine-tuningの2段階からなる。pre-trainingは大きな計算コストがかかるが、誰かが1回学習して公開すればよい。これらのリソースは、英語、中国語圏では急速に公開、共有されているが、日本では大幅に遅れており、研究者、研究室で役割分担・連携して、国を挙げて取り組む必要がある。

まさに今回の我々のようなチャレンジが不可欠だ。このような組織の垣根を越えたオープンコラボレーションが、Withコロナ時代にはマッチするとともに、より重要になると感じている。

「SNSによるデマ拡散」 問題の本質とは

「疑うタイミング」の創出が重要に

鳥海不二夫氏 Fujio Toriumi

東京大学大学院工学系研究科システム創成学専攻 准教授

聞き手：村山恵一氏 Keiichi Murayama

日本経済新聞社 コメンテーター

新型コロナウイルスの感染拡大を受け、世界的に問題視されたものの1つに「SNS（ソーシャルメディア）を通じたデマの拡散」がある（図1）。日本でもトイレ紙不足を招いたとされるなど、社会的に注目を集めた。この状況をどう認識、理解すべきなのか。「SNSとデマ」をめぐる問題の本質について、東京大学の鳥海不二夫准教授に聞いた。

「正しさ」より「面白さ」も背景に

——そもそもデマはどういう背景、経緯で出回のでしょうか。

鳥海 いくつかの理由があります。当初から騙してやろうというものはほとんど存在しないと言ってよく、どちらかというところ勘違いです。人間は同時に何かが起こると因果関係を求めたがる性質があります。その結果、勘違いをして喧伝するという流れです。情報が流布する原因は人々が慌てているからではありません。多くの人はほとんどの情報について、いちいち正しいかどうかは考えません。ほとんどの情報はとくに疑われることもなく拡散されるのです。それ自体は自然な流れだと思います。

面白いと思ったからその情報を拡散するというものもあります。あやしい内容かもしれないけれど、ほかの人に話題が提供できたり、友達同士で話が弾んだりする、という理由によるものです。「トイレ紙がなくなるらしいよ」のほうが、「今日もトイレ紙が棚にあったよ」というより面白い。情報の正しさを伝えるというより、楽しさを共有するという意味でデマが拡散するのです。それが2つ目の拡散要因です。

さらに、マスメディアがデマ的な情報を広めている事例も見

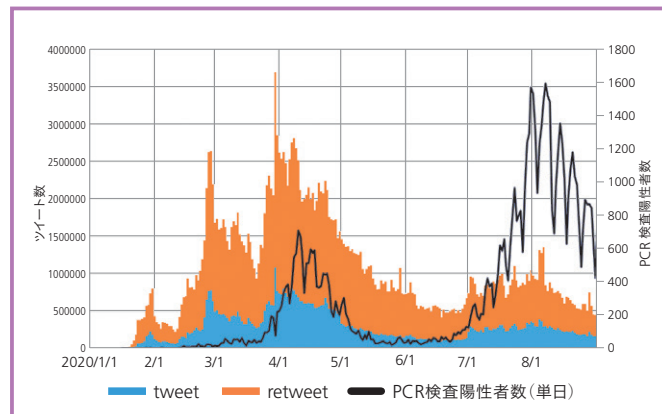


図1 コロナ禍での新型コロナに関するツイート数の時系列変化

られます。拡散につながる報道にはいくつかのパターンがありますが、新型コロナで特徴的だったのは、「SNS上でこんな話が盛り上がっている」と紹介する報道が多くなされ、そのなかにも、それほど拡散していない情報が含まれていたケースです。Twitterに書いている人は数人しかいないにもかかわらず、「こんなデマがTwitter上で拡散されている」と報じられるケースがありました。

——新型コロナに関連するデマの影響をどの程度と考えますか？

鳥海 いちばん話題になったトイレ紙不足に関するもののほか、「お湯を飲むと効果がある」「イソジンが新型コロナに効く」というものがありました。ただデマはありましたが、それが拡散したかというところでしょうか。正直なところ拡散したデマは多くないと思います。それほど広がっていない印象があります。

もっとも、どのくらいの量になれば拡散したというのは難しいところです。たとえばトイレ紙に関するデマを詳細に分析してみると、「トイレ紙がなくなる」「な



2004年、東京工業大学大学院理工学研究科機械制御システム工学専攻博士課程修了（博士〈工学〉）、2012年より現職。計算社会科学、人工知能技術の社会応用などの研究に従事。「科学技術への顕著な貢献2018（ナイスステップな研究者）」に選出。

鳥海不二夫

くなるかもしれない」といったツイートを見た可能性のある人は、「それはデマだから注意しよう」と否定するツイートを見た人と比べると、約2%と非常に少なかった。デマしか見ていないという人の数はかなり限られると言っている。デマを拡散したと犯人扱いされた人がいましたが、データを見ると、その人が書いた内容は拡散していませんでした。リツイート数はわずか1か2。しかも自分自身によるもので、拡散はゼロと言っている状態です(図2)。

ではなぜ、トイレトペーパー不足が起こったのか。それは、「デマに騙された人が買いに走るのではないか。であれば自分たちも買っておこう」となった結果だと思われます。しかもトイレトペーパーを買ったのはTwitterのメインユーザーである10～30代ではなく、Twitterをあまり使っていないもっと上の世代です。これをSNS発のデマが原因と言ってよいかというと、かなり疑問です。

問題は拡散の「量」ではない

—— SNSでのデマ拡散で社会が混乱したとは受け止めない方がいいと？

鳥海 SNSのユーザーは若い世代に偏っています。新型コロナウイルスに関する情報をどこで見聞きしたかについてアンケートをとると、SNSで得た人は10～30代が多く、40代は4分の1くらい、60代は10%以下。日本の人口のボリュームゾーンは年齢が高い方にあるため、SNSの影響の方が大きいとすると、おかしいことになる。また、SNSをどのくらい信用しているかという話になると、若年層さえ信頼度は20%を下回る。彼らもTwitterを信用できるものだとは思っていないのです。

少なくとも日本では、デマがすぐく拡散して、皆が騙されるような状況にはなっていません。一方で、大規模に拡散していても危険なデマは存在します。科学否定のようなものはそれなりにあり、そういうデマは生命に関わることもあるので危険です。しかもこうした問題は、拡散した数が多い、少ないというように量で測ることはできません。たとえ拡散する数が少なくても致命的なことになり得る。新型コロナもそうですが、災害時、緊急時のデマは問題になる可能性が高いと言えます。

緊急時に生じる「情報発信は社会貢献」？

—— 緊急時と平常時ではデマの発生具合に違いがあるのでしょうか。

鳥海 新型コロナや災害などは、皆が同時に興味をもつ話題です。基本的にSNS上の人々はコミュニティを形成して、自分の興味のある情報を手に入れるために使っています。それぞれのエリアで活動し、楽しみのために使っていることが多い。普段は異なるコミュニティ間で共通の話題はそれほどないため、コミュニティをまたいで拡散する情報は多くありません。一方、新型コロナのような話になると、皆が同時に興味をもち、拡散しやすい土壌ができるのです。

また、緊急時になると、人々は重要そうに見える情報を広めることに意義を感じます。とくに災害時に多いのですが、情報

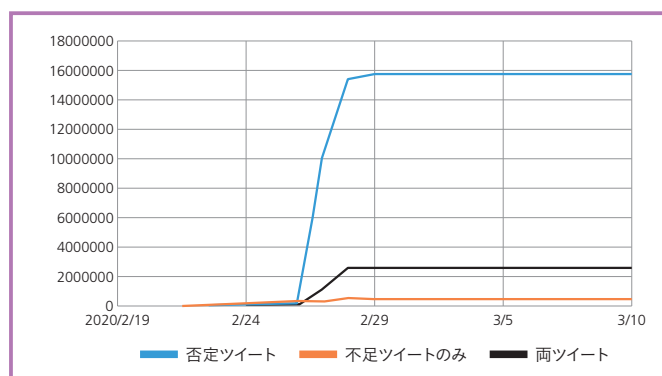


図2 | トイレトペーパー不足を否定するツイートと、不足を伝えるツイートの拡散状況

を広めることが社会的な活動だと思ってしまう。それをするだけで自分は社会的に良いことをしたという満足感を覚えることができるからです。つまり、自分自身はほとんど労力をかけずに社会的に貢献したと感じる。そうした行為が災害時などには多くなり、その結果、真偽が不確かなデマも拡散しやすくなります。—— こうした状況を踏まえて、SNSの活用は今後どうすべきでしょうか。

鳥海 これまで問題が起きていないから、今後もこれまで通りでいいという考え方は危険です。かといって情報のすべてを疑うのは難しいし、基本的には性善説に立つほうが世の中は回りやすいでしょう。ただ、情報について「疑うポイント」はいくつかあると思います。「この人が発信源ならあやしい」「この人のリツイートには偏りがある」などといった指標をもつことです。デマはポジショントークの場合も少なくありません。それを広めているのが誰なのかは確認した方がいいと思います。

つまり、情報の出どころや拡散者などに関する、いわゆるメタ情報が重要です。発信源となったメディアが信頼に値するのかどうかなども判断の手がかりになる。情報に接したとき、「疑わしいな」と思うことさえできれば、手を打つことも可能でしょう。その観点からメタ情報の提供によって、そうした「疑うタイミング」をつくることができればいいと思います。バックグラウンドの情報があれば、情報の確からしさを調べたり、ほかに違う意見があるかもしれないと気づいたりすることができるはずです。

インタビュアーからのひとこと

米Twitterの創業者ジャック・ドーシー最高経営責任者(CEO)と話したとき、Twitterの利用のされ方は「当初の想定を超えている」と言っていた。情報収集やコミュニケーションの道具として使いこなす人々の姿は「感動的」と評価は前向きだった。同時に、誤情報や偽情報、誹謗中傷問題への対応も怠れないと付け加えた。SNSが秘めるポテンシャルを私たちはまだ十分に引き出せていない。知恵を出す余地がたくさんある。

村山恵一

1992年日本経済新聞社入社。産業部で情報通信・エレクトロニクス、自動車、医療、金融などを担当。2004～05年米ハーバード大学留学。2005～09年米シリコンバレー支局。2012年編集委員。2015年論説委員を兼務。2017年からコメンテーター。担当はIT、スタートアップ。



ビッグデータ解析で コロナ社会の課題解決を

外出自粛率や人の接触頻度を「見える化」する

水野貴之 Takayuki Mizuno

国立情報学研究所 情報社会相関研究系 准教授
総合研究大学院大学 複合科学研究科 准教授

聞き手：山本佳世子氏 Kayoko Yamamoto

日刊工業新聞社 論説委員兼編集委員

情報学や物理学で培ったビッグデータ解析や大規模シミュレーションの知見を活かして、経済や社会の課題解決に挑む NII の水野貴之准教授。「計算社会科学」が専門のデータサイエンティストとして、携帯電話キャリアや全地球測位システム（GPS）などのビッグデータを解析し、新型コロナウイルス感染症対策に役立てるべく、積極的に情報発信をしている。さらに、AI（人工知能）を使って、時々刻々と変わる感染状況や、国によって異なる感染の広がりを捉えようと奮闘中だ。

データ解析結果を政策立案に役立てるために

—— 携帯電話から得た位置情報や購買履歴などのビッグデータを解析し、人の流れやビジネス活動をリアルタイムに可視化する研究が、ここ数年とくに注目されています。新型コロナウイルス対策への活用にも期待が集まっていますね。

水野 データを集めるベンダーは携帯キャリアの民間企業、データ解析の結果をもとに政策を決めるのは国や地方自治体、その間にいて、ビッグデータから必要なデータを加工して意味のある有用な情報を抽出する役割を担うのがデータサイエンティストです。コロナ禍では、このデータサイエンティストが不在の状況から新型コロナ対策に向けたデータ活用がスタートしたため、最初は混乱がありました。

データ解析自体は他分野でデータを扱う人にも可能ですが、正しい手法で解析しなければ、恣意的な情報やバイアスのかった情報が公表されてしまう危うい面があります。データに

ついてはそれぞれ癖があり、とくに公的統計に活用する際には注意が必要です。たとえば携帯キャリアの場合、NTTドコモなら、楽天モバイルよりユーザーの年齢層は高いし、利用者の多い地域も異なります。データ解析を正当に行うには、データや解析手法を熟知したデータサイエン

ティストが欠かせません。さらに、情報学の知見をもつ NII の研究者などが率先して関わらなくてはならない。そう判断して、私たちは3月半ばから感染症の研究者とも連携しながら、新型コロナ対策の研究を始めました。

外出自粛は「流入」ではなく「流出」を見る

—— 当初の新型コロナ対策で、政府が接触8割削減を掲げて外出自粛を求めたところ、都心の繁華街などの人出は減少しましたが、逆に近隣の商店街や公園の人出は増えてしまいました。

水野 これは、注目すべきデータ群を見誤ったのが原因です。外出自粛のためには、商業圏への人の「流入」ではなく、人々が住居地の外に出ていく「流出」を抑えることが重要だったのです。

そこで、私たち NII とキャノングローバル戦略研究所のグループは、この観点から研究に取り組みました。ドコモの携帯電話約8000万台の基地局情報から、推定されたリアルタイム人口分布を利用して、住宅地からの「外出者数（夜間人口－昼間人口）」を見積もり、平常時の外出者数に対する、新型コロナ下の外出者数の割合を使って自粛率を導きました。具体的には、各市町村の住民がコロナ以前に比べて外出をどの程度控えているかを、「 $1 - (\text{ある日の外出者数}) / (\text{平常時の外出者数})$ 」という形で見える化しています（図1）。このデータを NII のホームページで公表したところ、NHK が数カ月間にわたって報道し続けたことで、各人の行動に注意を促すことができました。また、自粛率と感染抑制の相関や、年代や性別によって活動の仕方が違うといったことも確認できました。

実は一般には公表していないデータとして、私たちは500m四方の区画（メッシュ）での流出も把握しています。これによ



水野貴之

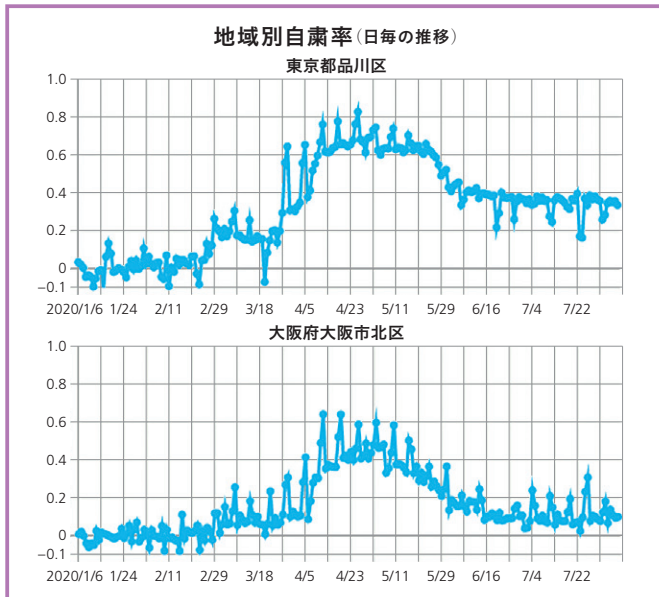


図1 | 流動人口ビッグデータによる地域民の自粛率の見える化。都道府県や地域別に自粛率を検索することができる。

り、自粛率の数値と、地域の保育園や町工場、商店街などの事情を関連づけることができます。県や、東京都なら市区町村レベルの行政が、このデータを活用して効果的な自粛要請を行えるようにもなりました。

なお、ピーク時の東京都内では自粛率6～8割減を達成し、8月半ば現在でも、東京の住宅地からの流出は2割ほど減ったままです。これはテレワークの影響でしょう。現在この解析結果は、経済活動との関連を把握する際にも活用されています。

目的が接触頻度の減少なら、何を制御すべきか

—— その後、政府の施策は一律の自粛要請ではなく、ソーシャルディスタンスをとった上での社会活動へと変わってきましたね。

水野 そこで私たちは、人から人への感染につながる「接触頻度」に注目し、GPS位置情報のビッグデータを活用して、人口分布の高解像度化に取り組みました。これには、ドコモの基地局から収集した約8000万人分の500m×500m精度の位置情報データと、許諾のとれた携帯電話使用者から集めた約20万人分の数m精度の位置情報データ、つまり「精度は低い为数の多いデータ」と、「精度は高いが数の少ないデータ」の2つを組み合わせました。

さらに、GPSでは2次元情報しかわかりませんが、高いビルの場合、各階に人がいれば実際には密ではないこと、人口密度が高くて車などで移動していれば問題ないこと、さらに人の向きなどの条件も加味しました。問題なのは15分以上、近い距離で顔と顔が近づいている状況です。これらの要素を考慮して解析した結果、単位時間あたりの人口密度の変化と、一人あたりの接触人数の変化が推定できるようになり、現実に近い接触頻度を導くことができるようになりました(図2)。この解析結果をもとに、「人の接触を8割減とする」には、「人出を8割減まで減らさなくても、6割5分程度の減少でよい」ということを明らかにしたのです。

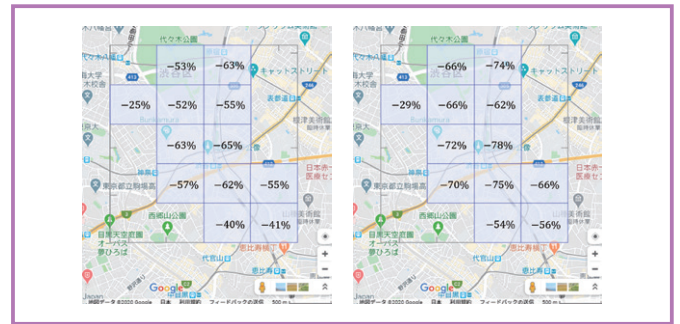


図2 | 渋谷周辺の平日(4月24日)の昼間のデータ(1月17日との比較)。500m四方区画の人口の変化率(左)と接触頻度の変化率(右=人口密度効果 $q=0.68$)を算出した。接触頻度を見れば、より現実に近い密の状態を知ることができる。これを基に、地域ごとに細やかな対策を打つことが可能になる。

AIでファクターXを見つけ出したい

—— いま、感染症の流行を予測するモデルの確立が求められています。自粛率や人の分布などさまざまなファクター(要因)がありますが、何が流行の抑制に効くのか、導き出せるのでしょうか?

水野 こういったモデルを導くには大きく2つの手法があります。1つは古典的な物理モデルをベースに、経済学や疫学などの理論から、因果関係のよく知られた重要なファクターのみを使って、感染流行を記述するモデルを組み立てます。この場合、どのファクターが、どのようにどの程度感染に効いているかは理解しやすいのですが、ファクターの数が少ないため、精度はあまり高くありません。一方、機械学習によるAIでは、大量のファクターを使って、高い精度で感染を予測することができます。中身がブラックボックスになっているため、どのファクターがどう効いているのかは人には理解できませんが、思いがけない未知のファクターを見つけ出すことには役立つはずです。

たとえば、元気な高齢者が昼間に営業しているスナックでカラオケを楽しむ、そこで感染が広がるという現象は、意外なものでした。こういった、その地域ならではの気づきにくいファクターを、AIなら見つけられるでしょう。日本を含むアジアと、欧米の感染者の広がり方の違いはなぜなのか、何がファクターXとして効いているのか、AIで探し出したいと考えています。これは、新型コロナに限った話ではなく、世界で起っているさまざまな現象を捉えるのにも役立つでしょう。

(写真=佐藤祐介)

インタビューからのひとこと

ビッグデータ解析でもアンケートでも、研究者は解析結果の公表時に強いメッセージを出そうとはしない。影響あるファクターの研究まで進んでいないうえ、専門外の領域では発言しないという、慎重さゆえだと理解していた。これに対して水野准教授は「活用したモデルには限界があり、必ず誤差があるからだ」と説明する。科学者として誠実な態度を示しつつ、社会の複雑な課題に積極果敢に取り組む水野准教授とNIIの活動に期待したい。

山本佳世子

日刊工業新聞社論説委員兼編集委員 1990年東京工業大学大学院修士課程修了後、日刊工業新聞社入社。科学技術、化学業界の担当を経て、「大学・産学連携」の専門記者で20年弱。同分野の研究で東京農工大学にて博士(学術)。文部科学省記者クラブ常駐。



教育が変わる時

安浦寛人

Hiroto Yasuura

九州大学 理事・副学長

それは、突然やってきた。昨年末に中国武漢で始まった COVID-19 の感染拡大は、2 月後半から我が国や欧米に広がり、3 月以降は世界全体を覆うパンデミックとなった。8 月末の時点で、全世界で 2500 万人が感染し、85 万人以上の犠牲者を出している。

春、花々は咲き誇り、山々は新しい芽吹きに覆われ、海や川は夏の装いになって行く。見慣れた街並みや風景も季節の変化を刻んで行くのに、人々の生活は、以前とは一変した。祭りやイベントは中止となり、通学や通勤の機会も減り、繁華街の賑わいも消えた。大学では、卒業式や入学式の式典もなく、突然、遠隔授業だけの開講と

なった。物理的な社会の構成はほとんど維持されたままで、我々は唐突にサイバー空間を中心とする世界での活動を強制されるようになった。

戦争や自然災害による社会の非常事態を想定していた人は多いであろうが、このような形での社会活動の変化を想定していた人は少ないであろう。しかし、見方を変えれば、これまで CPS (Cyber Physical System) の構築を進めてきた成果がうまく機能し、社会活動もある程度維持されていると見ることもできる。特に、大学においては、国内外の多くの大学が遠隔授業を実施して、次世代の育成である「教育」を止めずに継続できていることは、将来の歴史家もある程度評価してくれるのではないだろうか。

4 年前に、この Essay に「情報技術は大学革命の砦」という一文を投稿したが、そこで夢見ていたことの一部が実現している。さらに、研究や大学の運営業務などについても、これを機に、デジタル化 (DX: Digital Transformation) を実現しようという機運が高まってきている。

特に、教育の DX は、世界各国がほぼ同時に同じような条件下に置かれ、さまざまなアイデアを生み出し、ツールを構築し、運用の基準やノウハウを蓄積している。我が国でも、NII と主な大学が中心となって「4 月からの大学等遠隔授業に関する取組状況共有サイバーシンポジウム」を 15 回 (9 月 4 日時点) にわたって開催し、大学から初等・中等教育までのさまざまな工夫や実践例を共有するとともに、海外の事情も紹介してきた。10 年後には、2020 年を機に「教育」のあり方が本質的に変わったと言われるようになる可能性も大きいと思う。これは、世界的な潮流であり、今後、初等・中等教育から高等教育、社会人教育までを含む大きな潮流になると予想する。このような歴史的転換期に、情報技術やそれを用いた教育基盤の改革に少しでも参画できていることを、幸いに思うこの頃である。

参考文献

安浦寛人:「情報技術は大学革命の砦」, Essay, NII Today No.72, <https://www.nii.ac.jp/today/72/7.html>

今後の予定

10月17日～18日 | 大学共同利用機関シンポジウム2020(出展) = オンライン開催。お申し込み、詳細は <https://ius.4kikou.org/>

10月下旬 | 第2回 SPARC Japan セミナー2020 = オンライン開催。詳細は <https://www.nii.ac.jp/sparc/event/>

11月1日～30日 | 第22回図書館総合展(出展) = オンライン開催。コアタイムあり。詳細は <https://www.nii.ac.jp/event/other/libraryfair/>

12月9日～11日 | 大学ICT推進協議会 (AXIES) 2020 年度年次大会(出展) = オンライン開催。詳細は <https://axies.jp/conf/conf2020/>

Information

2021年度 国立情報学研究所 公募型共同研究の募集を開始

公募する共同研究は以下の3種類で、募集期間は2020年10月1日～12月1日です。コロナ禍を乗り越えるための研究提案も募集します。若手研究者、女性研究者、遠方の大学等の研究者の皆様のご応募も歓迎します。

① 戦略研究公募型 ② 研究企画会合公募型 ③ 自由提案公募型

応募方法や応募資格等の詳細は、ウェブサイトをご覧ください。

<https://www.nii.ac.jp/research/collaboration/koubo/>

表紙の言葉

肺の CT 画像に見られるのは、COVID-19 肺炎に特徴的なすりガラス状の影。本来、金属製のロボットは CT 検査を受けることはできませんが、人間であれば COVID-19 らしさや肺の状態を確かめるのに役立ちます。この画像を教師データに、診断支援 AI の開発が進められています。

情報から知を紡ぎだす。

NII

国立情報学研究所ニュース [NII Today] 第89号 令和2年9月

発行 | 大学共同利用機関法人 情報・システム研究機構 国立情報学研究所
〒101-8430 東京都千代田区一ツ橋2丁目1番2号 学術総合センター

発行人 | 喜連川 優 編集長 | 佐藤一郎

表紙画 | 城谷俊也 編集 | 田井中麻都佳

制作 | 株式会社マツダオフィス / サイテック・コミュニケーションズ

本誌についてのお問い合わせ | 総務部企画課 広報チーム

TEL | 03-4212-2028 FAX | 03-4212-2150 e-mail | kouhou@nii.ac.jp

「NII Today」で
検索!



情報犬ビット
(NII キャラクター)

<https://www.nii.ac.jp/today/>